Text as Data

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

October 30th, 2014

Structured Topic Models

1) Task:
    - Examine how document attention, topic content varies $\rightsquigarrow$ over time, across authors, or with general set of covariates
2) Objective Function

$$f(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{W})$$

Where:
    - $\boldsymbol{W}$ condition on information in document $\rightsquigarrow$ other potential modifications to objective function. Meta-data.
    - $f(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{W})$ may encode additional information $\rightsquigarrow$ layers of clustering, layers of topics, etc
3) Optimization
    - EM, Variational Approximation, Gibbs Sampling, ...
4) Validation $\rightsquigarrow$ many of the same methods from clustering
    - Semantic, Convergent, Discriminant, Predictive, Hypothesis validity
    - How do we avoid the electric machine critique?

LDA Revisited

$$
\begin{aligned}
\boldsymbol{\theta}_k &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\pi}_i|\boldsymbol{\alpha} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
\boldsymbol{\tau}_{im}|\boldsymbol{\pi}_i &\sim \text{Multinomial}(1, \boldsymbol{\pi}_i) \\
x_{im}|\boldsymbol{\theta}_k, \tau_{imk} = 1 &\sim \text{Multinomial}(1, \boldsymbol{\theta}_k)
\end{aligned}
$$

LDA Revisited

$$
\begin{aligned}
\textbf{Unigram Model}_k &\sim \text{Dirichlet}(\mathbf{1}) \\
\textbf{Doc. Prop}_i &\sim \text{Dirichlet}(\textbf{Pop. Proportion}) \\
\textbf{Word Topic}_{im} &\sim \text{Multinomial}(1, \textbf{Doc. Prop}_i) \\
\text{Word}_{im} &\sim \text{Multinomial}(1, \textbf{Unigram Model}_k)
\end{aligned}
$$

A General Hierarchical Structure

LDA:

Pop. Proportion

A General Hierarchical Structure

LDA:

Pop. Proportion

Doc. Prop$_1$ **Doc. Prop**$_2$ $\cdots$ **Doc. Prop**$_N$

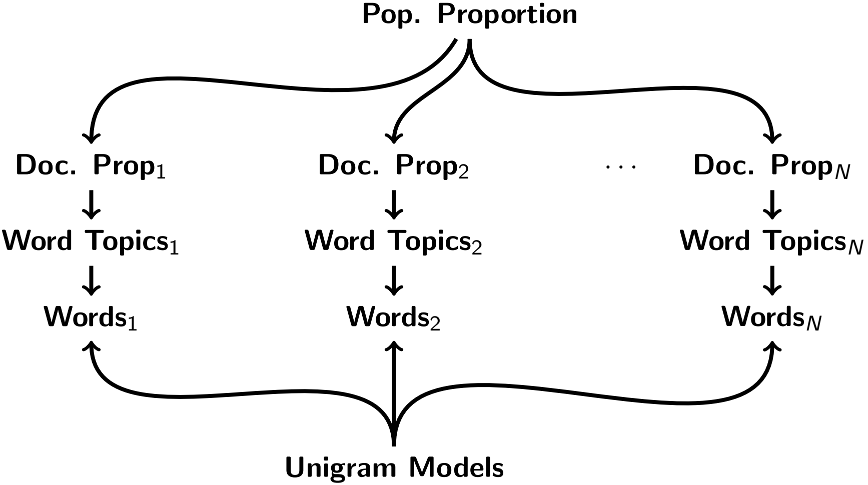
A General Hierarchical Structure

LDA:

$$
\begin{tikzcd}[column sep=small]
 & & \textbf{Pop. Proportion} \arrow[dll, bend right] \arrow[dl, bend right] \arrow[drr, bend left] & & \\
\textbf{Doc. Prop}_1 \arrow[d] & \textbf{Doc. Prop}_2 \arrow[d] & & \cdots & \textbf{Doc. Prop}_N \arrow[d] \\
\textbf{Word Topics}_1 & \textbf{Word Topics}_2 & & & \textbf{Word Topics}_N
\end{tikzcd}
$$

A General Hierarchical Structure

LDA:



A General Hierarchical Structure

Dynamic Topic Model (Quinn et al 2010)

Dynamic Prior Across Days

A General Hierarchical Structure

Dynamic Topic Model (Quinn et al 2010)

Dynamic Prior Across Days

Day Prop$_1$ **Day Prop**$_2$ $\cdots$ **Day Prop**$_N$

A General Hierarchical Structure

Dynamic Topic Model (Quinn et al 2010)

Dynamic Prior Across Days

Day Prop$_1$ → **Speech Topics$_1$**

Day Prop$_2$ → **Speech Topics$_2$**

$\cdots$

Day Prop$_N$ → **Speech Topics$_N$**

A General Hierarchical Structure

Dynamic Topic Model (Quinn et al 2010)

Dynamic Prior Across Days

Day Prop$_1$ → **Speech Topics$_1$** → **Speeches$_1$**

Day Prop$_2$ → **Speech Topics$_2$** → **Speeches$_2$**

$\cdots$

Day Prop$_N$ → **Speech Topics$_N$** → **Speeches$_N$**

Unigram Models

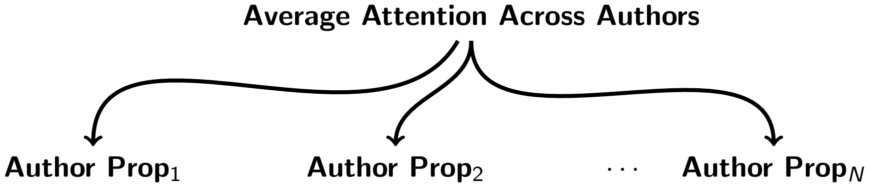
A General Hierarchical Structure

Expressed Agenda Model (Grimmer 2010)

Average Attention Across Authors

A General Hierarchical Structure

Expressed Agenda Model (Grimmer 2010)



A General Hierarchical Structure

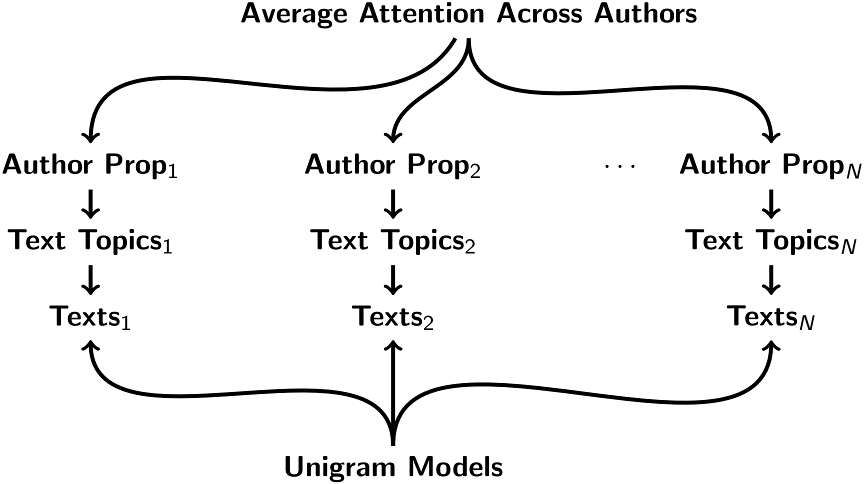
Expressed Agenda Model (Grimmer 2010)

Average Attention Across Authors

Author Prop$_1$	**Author Prop**$_2$	$\cdots$	**Author Prop**$_N$
$\downarrow$	$\downarrow$		$\downarrow$
Text Topics$_1$	**Text Topics**$_2$		**Text Topics**$_N$

A General Hierarchical Structure

Expressed Agenda Model (Grimmer 2010)



A General Hierarchical Structure

Structural Topic Model (Roberts, Stewart, Airoldi 2014)

$$g(\boldsymbol{W}, \gamma)$$

A General Hierarchical Structure

Structural Topic Model (Roberts, Stewart, Airoldi 2014)

$$g(\boldsymbol{W}, \gamma)$$

Strata Prop$_1$ **Strata Prop**$_2$ $\cdots$ **Strata Prop**$_N$

A General Hierarchical Structure

Structural Topic Model (Roberts, Stewart, Airoldi 2014)

$g(\boldsymbol{W}, \gamma)$

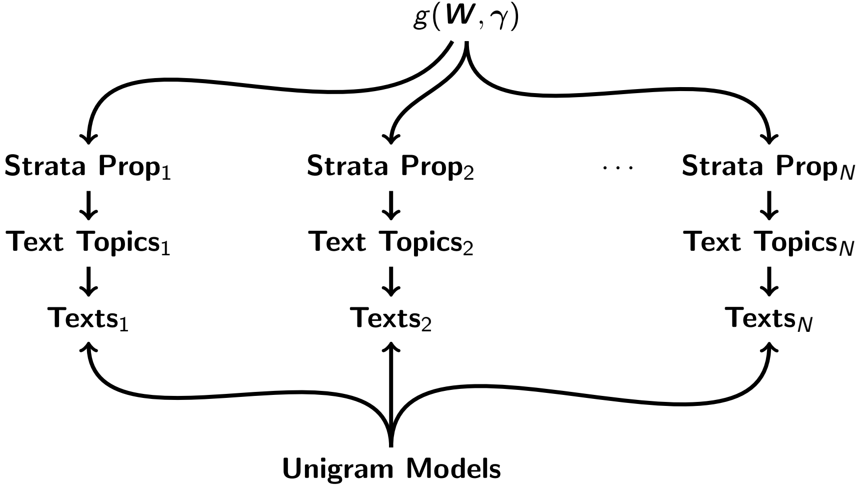
Strata Prop$_1$ **Strata Prop**$_2$ $\cdots$ **Strata Prop**$_N$

$\downarrow$ $\downarrow$ $\downarrow$

Text Topics$_1$ **Text Topics**$_2$ **Text Topics**$_N$

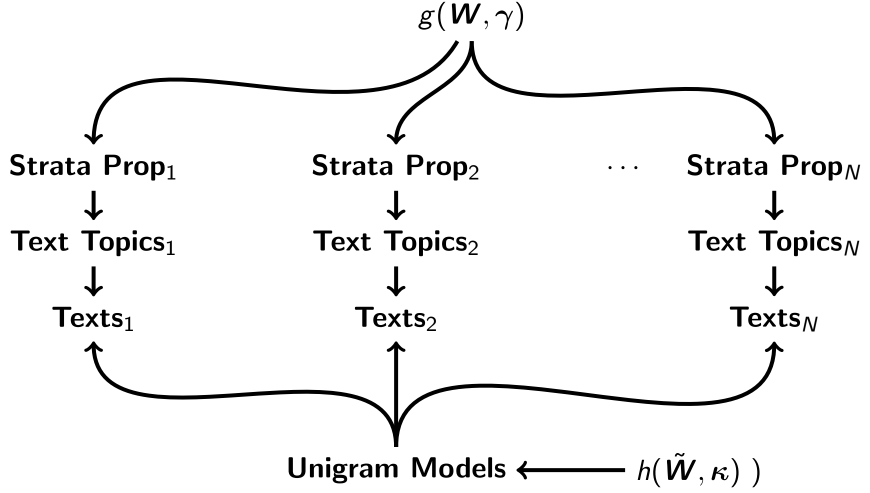
A General Hierarchical Structure

Structural Topic Model (Roberts, Stewart, Airoldi 2014)



A General Hierarchical Structure

Structural Topic Model (Roberts, Stewart, Airoldi 2014)



A General Hierarchical Structure

Conditioning on Unknown Covariates$\rightsquigarrow$ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)

Mixture of Top. Attn. Models

A General Hierarchical Structure

Conditioning on Unknown Covariates$\rightsquigarrow$ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)

Strata Indic. **Topic Attention Models**

A General Hierarchical Structure

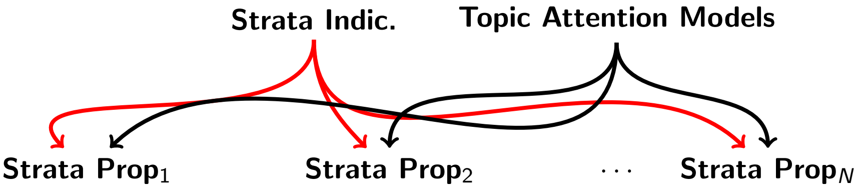
Conditioning on Unknown Covariates$\rightsquigarrow$ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)

Strata Indic. **Topic Attention Models**

Strata Prop$_1$ **Strata Prop**$_2$ $\cdots$ **Strata Prop**$_N$

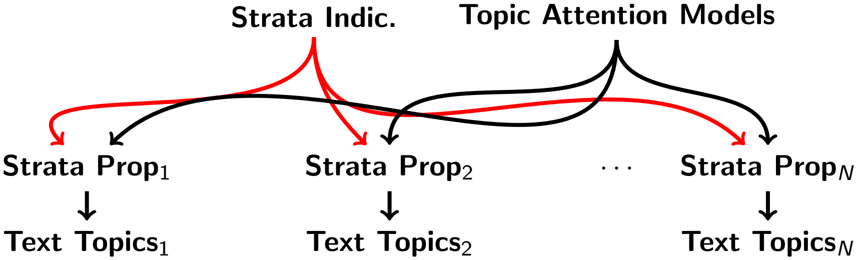
A General Hierarchical Structure

Conditioning on Unknown Covariates⇝ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)



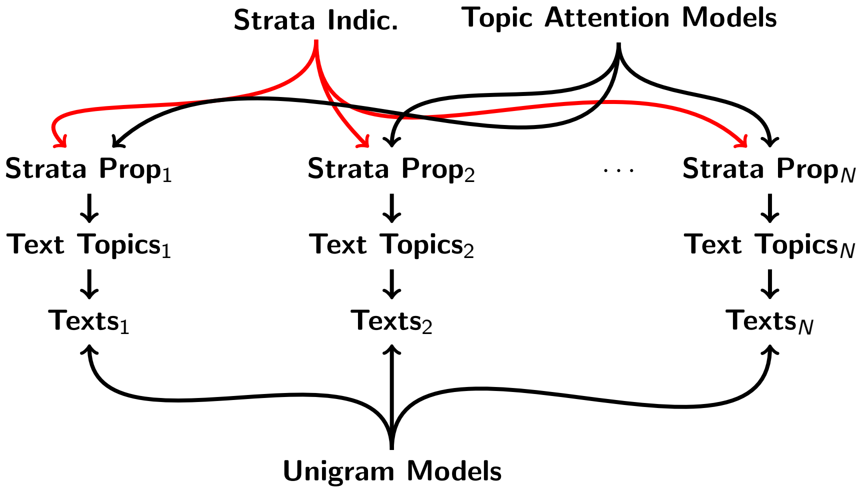
A General Hierarchical Structure

Conditioning on Unknown Covariates $\rightsquigarrow$ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)



A General Hierarchical Structure

Conditioning on Unknown Covariates$\rightsquigarrow$ levels of mixtures at proportions (Grimmer 2013; Wallach 2008)



A General Hierarchical Structure

Conditioning on Unknown Covariates for Topics $\rightsquigarrow$ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)

Population attention

A General Hierarchical Structure

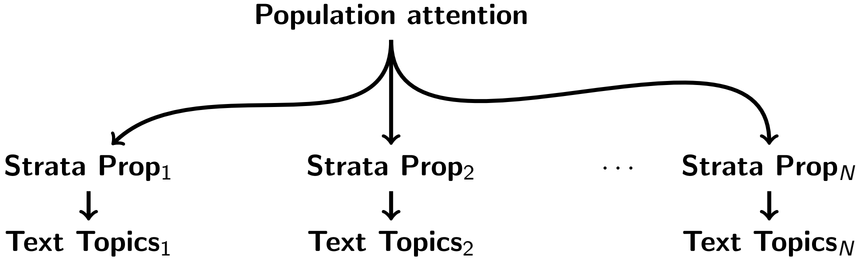
Conditioning on Unknown Covariates for Topics$\rightsquigarrow$ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)

Population attention

Strata Prop$_1$ **Strata Prop**$_2$ $\cdots$ **Strata Prop**$_N$

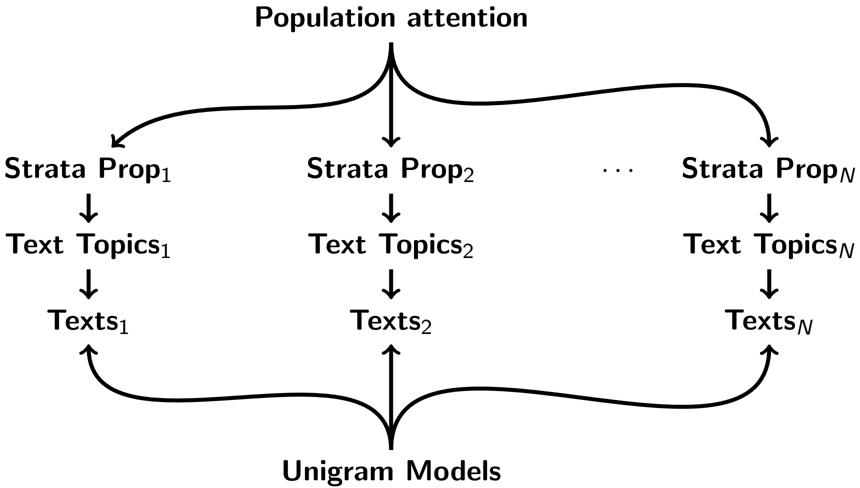
A General Hierarchical Structure

Conditioning on Unknown Covariates for Topics $\rightsquigarrow$ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)



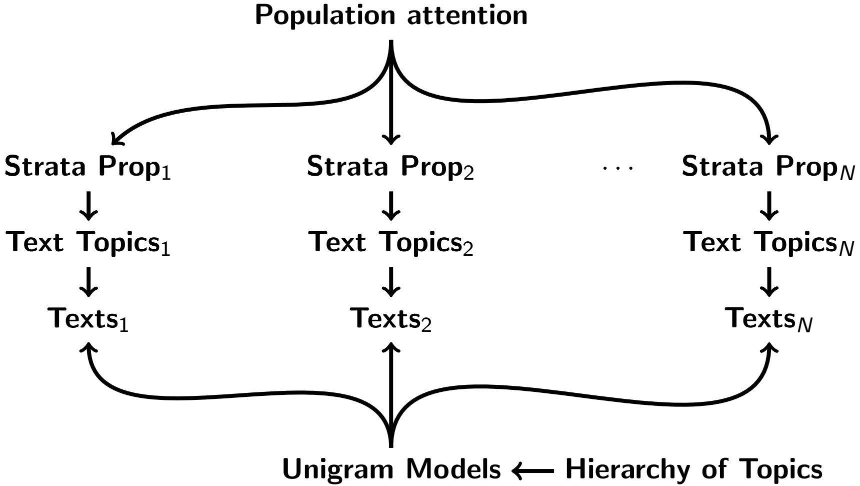
A General Hierarchical Structure

Conditioning on Unknown Covariates for Topics $\rightsquigarrow$ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)



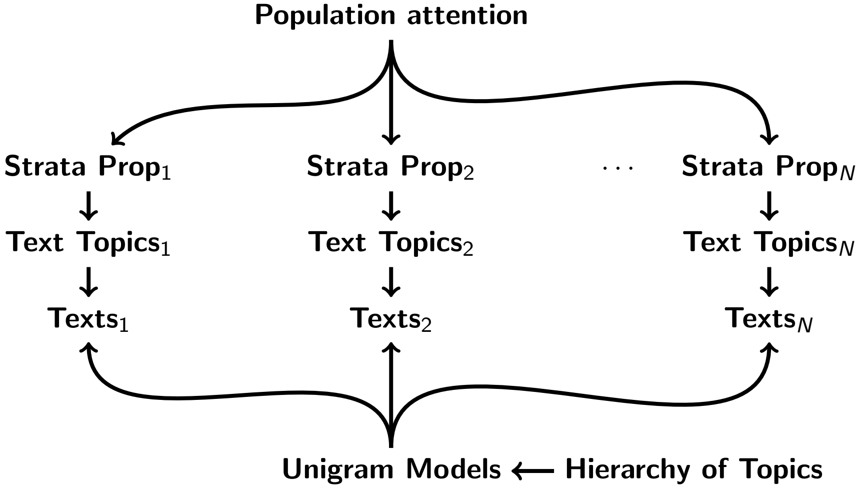
A General Hierarchical Structure

Conditioning on Unknown Covariates for Topics⇝ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)



A General Hierarchical Structure

Conditioning on Unknown Covariates for Topics$\rightsquigarrow$ hierarchy of topics (Li and McCallum 2006; Blaydes, Grimmer, and McQueen 2014)



Why Encode Structure in Extensions of LDA?

Why Encode Structure in Extensions of LDA?

- Substantive reasons

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis
  - Clear data generating process

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis
  - Clear data generating process
- Statistical reasons

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis
  - Clear data generating process
- Statistical reasons
  - Smoothing $\rightsquigarrow$ borrow information across groups intelligently

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis
  - Clear data generating process
- Statistical reasons
  - Smoothing $\rightsquigarrow$ borrow information across groups intelligently
  - Uncertainty $\rightsquigarrow$ potential for better uncertainty estimates

Why Encode Structure in Extensions of LDA?

- Substantive reasons
  - Additional structure corresponds to substantively interesting content
  - Avoids potential ad-hoc secondary analysis
  - Clear data generating process
- Statistical reasons
  - Smoothing $\rightsquigarrow$ borrow information across groups intelligently
  - Uncertainty $\rightsquigarrow$ potential for better uncertainty estimates
  - Improved topics $\rightsquigarrow$ small word conditions, structure could help

Plan for the Class

1) Discuss model with unknown covariates for strata proportions $\rightsquigarrow$ presentational style
2) Discuss model with hierarchy of topics $\rightsquigarrow$ mirrors genre

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:
Senators (representatives) regularly engage the public $\rightarrow$ presentational style
But we know little about this engagement

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:
Senators (representatives) regularly engage the public $\rightarrow$ presentational style
But we know little about this engagement
Why? Hard to Measure

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:
Senators (representatives) regularly engage the public $\rightarrow$ presentational style
But we know little about this engagement
Why? Hard to Measure
Describe model that facilitates estimation of presentational styles in Senate press releases

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:

Senators (representatives) regularly engage the public $\rightarrow$ presentational style

But we know little about this engagement

Why? Hard to Measure

Describe model that facilitates estimation of presentational styles in Senate press releases

- Characterize representation provided to constituents

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:

Senators (representatives) regularly engage the public $\rightarrow$ presentational style

But we know little about this engagement

Why? Hard to Measure

Describe model that facilitates estimation of presentational styles in Senate press releases

- Characterize representation provided to constituents
- Divide attention over a set of topics

Unknown Covariates for Issue Attention: Measuring Attention in Senate Press Releases

Substantive problem:
Senators (representatives) regularly engage the public $\rightarrow$ presentational style
But we know little about this engagement
Why? Hard to Measure
Describe model that facilitates estimation of presentational styles in Senate press releases

- Characterize representation provided to constituents
- Divide attention over a set of topics
- Given attention to topics, write press releases

Presentational Styles$\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Presentational Styles $\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Press release-level parameters (press release j from senator i in year t)

Presentational Styles $\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Press release-level parameters (press release j from senator i in year t)

- Assume: Each press release j assigned to one topic.
- Let $\boldsymbol{\tau}_{ijt}$ indicate press release j's topic.

Presentational Styles$\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Press release-level parameters (press release j from senator i in year t)

- Assume: Each press release j assigned to one topic.
- Let $\boldsymbol{\tau}_{ijt}$ indicate press release j's topic.

$$\boldsymbol{\tau}_{ijt} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{it})$$

Presentational Styles$\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Press release-level parameters (press release j from senator i in year t)

- Assume: Each press release j assigned to one topic.
- Let $\boldsymbol{\tau}_{ijt}$ indicate press release j's topic.

$$\boldsymbol{\tau}_{ijt} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{it})$$

- Conditional on topic, draw document's content.

Presentational Styles $\rightsquigarrow$ Objective Function

- $\pi_{itk} \equiv$ Attention senator i allocates to issue k in year t
- $\pi_{itk} \equiv$ Probability press release is about issue k
- $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it44})$

Press release-level parameters (press release j from senator i in year t)

- Assume: Each press release j assigned to one topic.
- Let $\boldsymbol{\tau}_{ijt}$ indicate press release j's topic.

$$\boldsymbol{\tau}_{ijt} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{it})$$

- Conditional on topic, draw document's content.
- If $\tau_{ijtk} = 1$ then

$$\boldsymbol{x}_{ijt} \sim \text{Multinomial}(n_{ijt}, \boldsymbol{\theta}_k).$$

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\boldsymbol{\sigma}_{it} \sim \text{Multinomial}(1, \boldsymbol{\beta}).$$

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\begin{aligned} \boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}). \\ \boldsymbol{\pi}_{it}|\sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \end{aligned}$$

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\begin{aligned} \boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}). \\ \boldsymbol{\pi}_{it} | \sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\ \alpha_{ks} &\sim \text{Gamma}(0.25, 1) \end{aligned}$$

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\begin{aligned}
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}). \\
\boldsymbol{\pi}_{it} | \sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1)
\end{aligned}$$

Other priors:

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\begin{aligned} \boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}). \\ \boldsymbol{\pi}_{it} | \sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\ \alpha_{ks} &\sim \text{Gamma}(0.25, 1) \end{aligned}$$

Other priors:

$$\boldsymbol{\theta}_k \sim \text{Multinomial}(\boldsymbol{\lambda})$$

Priors

Each $\boldsymbol{\pi}_{it}$ is a draw from one-of-S styles$\rightsquigarrow$ mixture of Dirichlet distributions .

$$\begin{aligned}
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}). \\
\boldsymbol{\pi}_{it}|\sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1)
\end{aligned}$$

Other priors:

$$\begin{aligned}
\boldsymbol{\theta}_k &\sim \text{Multinomial}(\boldsymbol{\lambda}) \\
\boldsymbol{\beta} &\sim \text{Multinomial}(\mathbf{1})
\end{aligned}$$

Presentational Styles $\rightsquigarrow$ Objective Function

Presentational Styles$\rightsquigarrow$ Objective Function

$$\begin{aligned} \boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\ \boldsymbol{\theta}_k &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \\ \alpha_{ks} &\sim \text{Gamma}(0.25, 1) \end{aligned}$$

Presentational Styles$\rightsquigarrow$ Objective Function

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}_k &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1) \\
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta})
\end{aligned}
$$

Presentational Styles$\rightsquigarrow$ Objective Function

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}_k &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1) \\
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}) \\
\boldsymbol{\pi}_{it}|\sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s)
\end{aligned}
$$

Presentational Styles $\rightsquigarrow$ Objective Function

$$
\begin{aligned}
\boldsymbol{\beta} & \sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}_k & \sim \text{Dirichlet}(\boldsymbol{\lambda}) \\
\alpha_{ks} & \sim \text{Gamma}(0.25, 1) \\
\boldsymbol{\sigma}_{it} & \sim \text{Multinomial}(1, \boldsymbol{\beta}) \\
\boldsymbol{\pi}_{it} | \sigma_{its} = 1, \boldsymbol{\alpha}_s & \sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\
\boldsymbol{\tau}_{ijt} | \boldsymbol{\pi}_{it} & \sim \text{Multinomial}(1, \boldsymbol{\pi}_{it})
\end{aligned}
$$

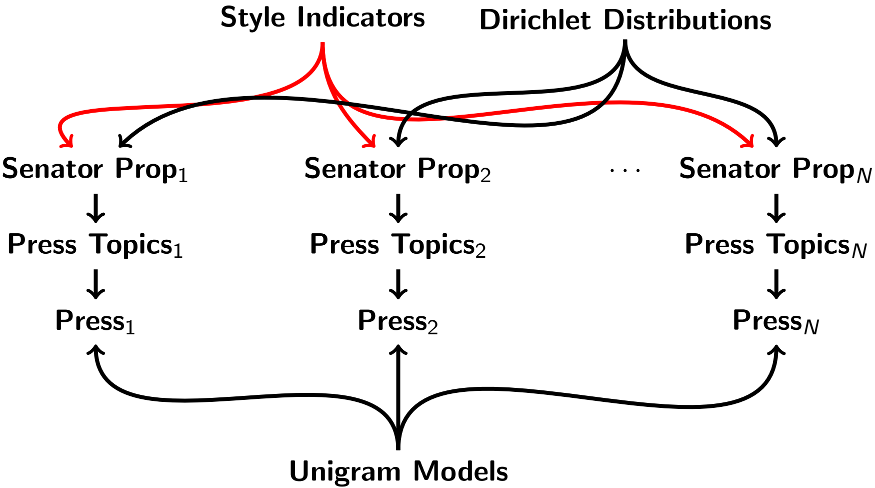
Presentational Styles $\rightsquigarrow$ Objective Function

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}_k &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1) \\
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}) \\
\boldsymbol{\pi}_{it}|\sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\
\boldsymbol{\tau}_{ijt}|\boldsymbol{\pi}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\pi}_{it}) \\
\boldsymbol{x}_{ijt}|\tau_{ijtk} = 1, \boldsymbol{\theta}_k &\sim \text{Multinomial}(n_{ijt}, \boldsymbol{\theta}_k)
\end{aligned}
$$

Presentational Styles$\rightsquigarrow$ Objective Function

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}(\mathbf{1}) \\
\boldsymbol{\theta}_k &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \\
\alpha_{ks} &\sim \text{Gamma}(0.25, 1) \\
\boldsymbol{\sigma}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\beta}) \\
\boldsymbol{\pi}_{it} | \sigma_{its} = 1, \boldsymbol{\alpha}_s &\sim \text{Dirichlet}(\boldsymbol{\alpha}_s) \\
\boldsymbol{\tau}_{ijt} | \boldsymbol{\pi}_{it} &\sim \text{Multinomial}(1, \boldsymbol{\pi}_{it}) \\
\boldsymbol{x}_{ijt} | \tau_{ijtk} = 1, \boldsymbol{\theta}_k &\sim \text{Multinomial}(n_{ijt}, \boldsymbol{\theta}_k)
\end{aligned}
$$

Mixture of Styles, Mixture of Topics



Posterior:

$$
\begin{aligned}
p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{X}) \quad \propto \quad & \prod_{k=1}^{K} \prod_{s=1}^{S} \frac{\exp(-\frac{\alpha_{ks}}{1/4})}{1/4} \times \frac{\Gamma(\sum_{w=1}^{W} \lambda_w)}{\prod_{w=1}^{W} \Gamma(\lambda_w)} \prod_{w=1}^{W} \theta_{k,w}^{\lambda_w - 1} \times \\
& \prod_{i=1}^{N} \prod_{t=2005}^{2007} \prod_{s=1}^{S} \left[\beta_s \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ks})}{\prod_{k=1}^{K} \Gamma(\alpha_{ks})} \prod_{k=1}^{K} \pi_{itk}^{\alpha_{ks}-1} \prod_{j=1}^{D_{it}} \prod_{k=1}^{K} \left[\pi_{itk} \prod_{w=1}^{W} \theta_{kw}^{x_{ijtw}} \right]^{\tau_{ijtk}} \right]^{\sigma_{its}}
\end{aligned}
$$

Posterior:

$$
\begin{aligned}
p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{X}) \quad \propto \quad & \prod_{k=1}^{K} \prod_{s=1}^{S} \frac{\exp(-\frac{\alpha_{ks}}{1/4})}{1/4} \times \frac{\Gamma(\sum_{w=1}^{W} \lambda_w)}{\prod_{w=1}^{W} \Gamma(\lambda_w)} \prod_{w=1}^{W} \theta_{k,w}^{\lambda_w - 1} \times \\
& \prod_{i=1}^{N} \prod_{t=2005}^{2007} \prod_{s=1}^{S} \left[\beta_s \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ks})}{\prod_{k=1}^{K} \Gamma(\alpha_{ks})} \prod_{k=1}^{K} \pi_{itk}^{\alpha_{ks} - 1} \prod_{j=1}^{D_{it}} \prod_{k=1}^{K} \left[\pi_{itk} \prod_{w=1}^{W} \theta_{kw}^{x_{ijtw}} \right]^{\tau_{ijtk}} \right]^{\sigma_{its}}
\end{aligned}
$$

1) Estimate with Variational Approximation

Posterior:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{X}) \propto \prod_{k=1}^{K} \prod_{s=1}^{S} \frac{\exp(-\frac{\alpha_{ks}}{1/4})}{1/4} \times \frac{\Gamma(\sum_{w=1}^{W} \lambda_w)}{\prod_{w=1}^{W} \Gamma(\lambda_w)} \prod_{w=1}^{W} \theta_{k,w}^{\lambda_w - 1} \times$$
$$\prod_{i=1}^{N} \prod_{t=2005}^{2007} \prod_{s=1}^{S} \left[\beta_s \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ks})}{\prod_{k=1}^{K} \Gamma(\alpha_{ks})} \prod_{k=1}^{K} \pi_{itk}^{\alpha_{ks} - 1} \prod_{j=1}^{D_{it}} \prod_{k=1}^{K} \left[\pi_{itk} \prod_{w=1}^{W} \theta_{kw}^{x_{ijtw}} \right]^{\tau_{ijtk}} \right]^{\sigma_{its}}$$

1) Estimate with Variational Approximation
2) Determining number of clusters at top? (Grimmer, Shorey, Wallach, and Zlotnick, In Progress)

Posterior:

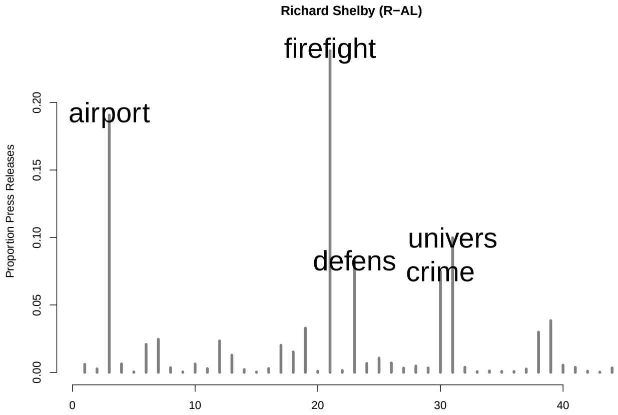
$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau}|\boldsymbol{X}) \propto \prod_{k=1}^{K}\prod_{s=1}^{S}\frac{\exp(-\frac{\alpha_{ks}}{1/4})}{1/4} \times \frac{\Gamma(\sum_{w=1}^{W}\lambda_w)}{\prod_{w=1}^{W}\Gamma(\lambda_w)}\prod_{w=1}^{W}\theta_{k,w}^{\lambda_w - 1} \times$$
$$\prod_{i=1}^{N}\prod_{t=2005}^{2007}\prod_{s=1}^{S}\left[\beta_s\frac{\Gamma(\sum_{k=1}^{K}\alpha_{ks})}{\prod_{k=1}^{K}\Gamma(\alpha_{ks})}\prod_{k=1}^{K}\pi_{itk}^{\alpha_{ks}-1}\prod_{j=1}^{D_{it}}\prod_{k=1}^{K}\left[\pi_{itk}\prod_{w=1}^{W}\theta_{kw}^{x_{ijtw}}\right]^{\tau_{ijtk}}\right]^{\sigma_{its}}$$

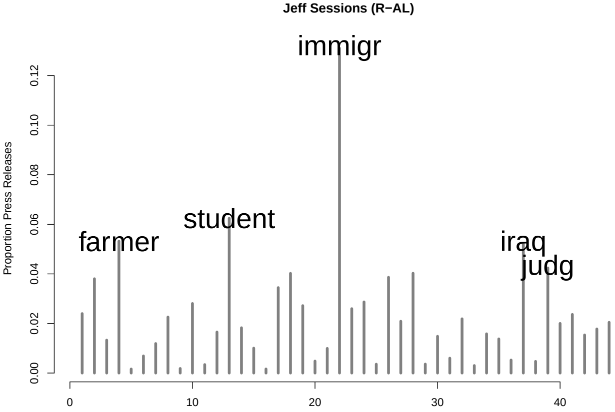
1) Estimate with Variational Approximation
2) Determining number of clusters at top? (Grimmer, Shorey, Wallach, and Zlotnick, In Progress)
   - Non-parametric model $\rightsquigarrow$ statistical selection

Posterior:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{X}) \propto \prod_{k=1}^{K} \prod_{s=1}^{S} \frac{\exp(-\frac{\alpha_{ks}}{1/4})}{1/4} \times \frac{\Gamma(\sum_{w=1}^{W} \lambda_w)}{\prod_{w=1}^{W} \Gamma(\lambda_w)} \prod_{w=1}^{W} \theta_{k,w}^{\lambda_w - 1} \times$$
$$\prod_{i=1}^{N} \prod_{t=2005}^{2007} \prod_{s=1}^{S} \left[\beta_s \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ks})}{\prod_{k=1}^{K} \Gamma(\alpha_{ks})} \prod_{k=1}^{K} \pi_{itk}^{\alpha_{ks} - 1} \prod_{j=1}^{D_{it}} \prod_{k=1}^{K} \left[\pi_{itk} \prod_{w=1}^{W} \theta_{kw}^{x_{ijtw}} \right]^{\tau_{ijtk}} \right]^{\sigma_{its}}$$

1) Estimate with Variational Approximation
2) Determining number of clusters at top? (Grimmer, Shorey, Wallach, and Zlotnick, In Progress)
    - Non-parametric model $\rightsquigarrow$ statistical selection
    - Experiments/Coding Exercises to assess


Richard Shelby (R−AL)
firefight
airport
univers
defens
crime
Proportion Press Releases
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40


Jeff Sessions (R−AL)
immigr
student
farmer
iraq
judg
Proportion Press Releases
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
40

Notions of validity: From Quinn, Monroe, et al (2010)

- Semantic Validity: All categories are coherent and meaningful
- Convergent Construct Validity: Measures concur with existing measures in critical details.
- Discriminant Construct Validity: Measures differ from existing measures in productive ways.
- Predictive Measure: Measures from the model corresponds to external events in expected ways.
- Hypothesis Validity: Measures generated from the model can be used to test substantive hypotheses.

To establish utility of new measures, demonstrate variety of validations
None of these validations are performed using a canned statistic
All: require substantive knowledge on areas (and what we expect!) [

Home Style Measures, Semantic Validity

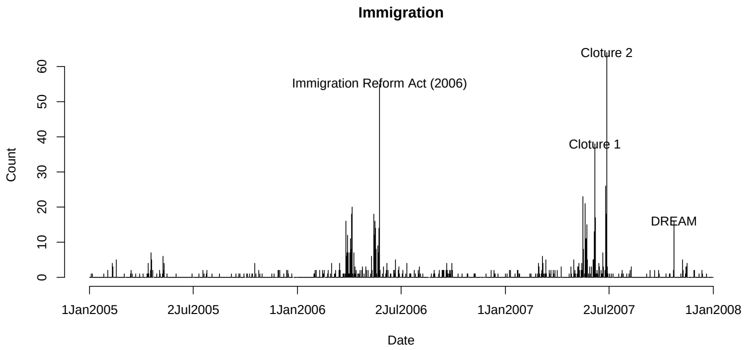
Must: Demonstrate to reader that topics are coherent and semantically meaningful

Description	Stems	%
Honorary	honor,prayer,rememb,fund,tribut	5.0
Transp. Grants	airport,transport,announc,urban,hud	4.8
Iraq	iraq,iraqi,troop,war,sectarian	4.7
DHS Policy	homeland,port,terrorist,dh,fema	4.1
Judicial Nom.	judg,court,suprem,nomin,nomine	3.8
Fire Dept. Grant	firefight,homeland,afgp,award,equip	3.7

How: examples in text are also useful.

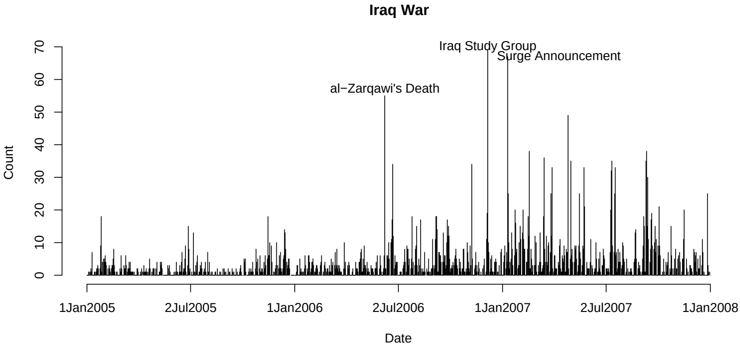
Home Style Measures, Convergent Validity

Over time variation



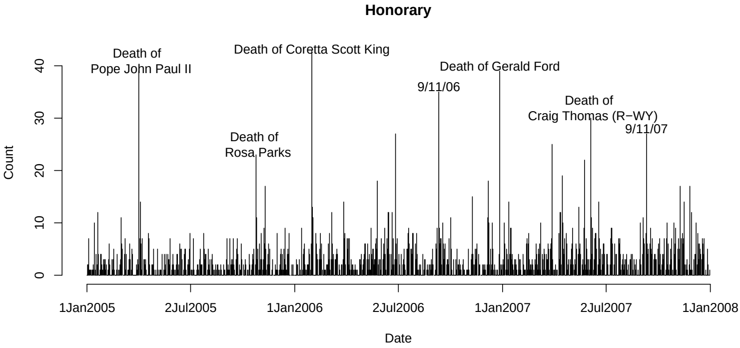
Home Style Measures, Convergent Validity

Over time variation



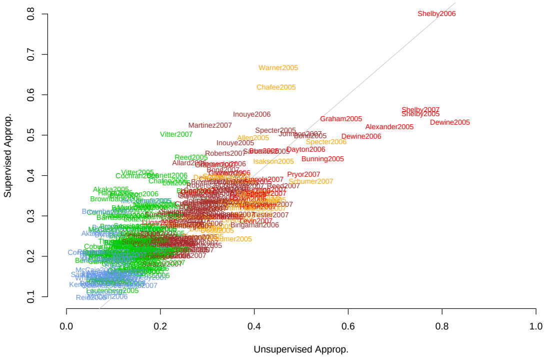
Home Style Measures, Convergent Validity

Over time variation

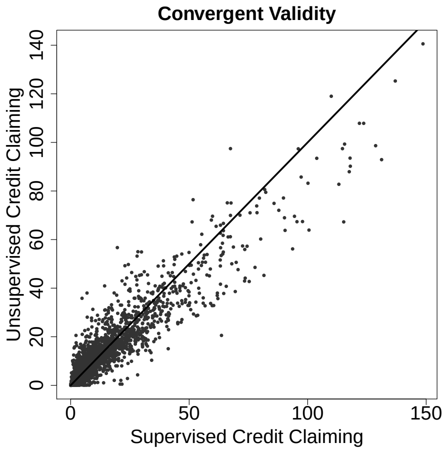


Home Style Measures, Convergent Validity

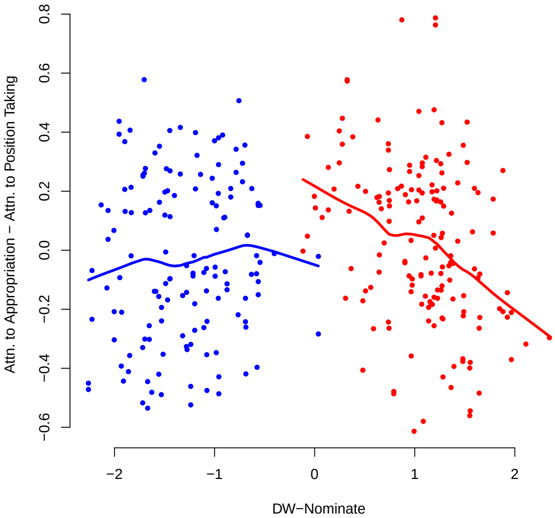
Supervised/Unsupervised Convergence



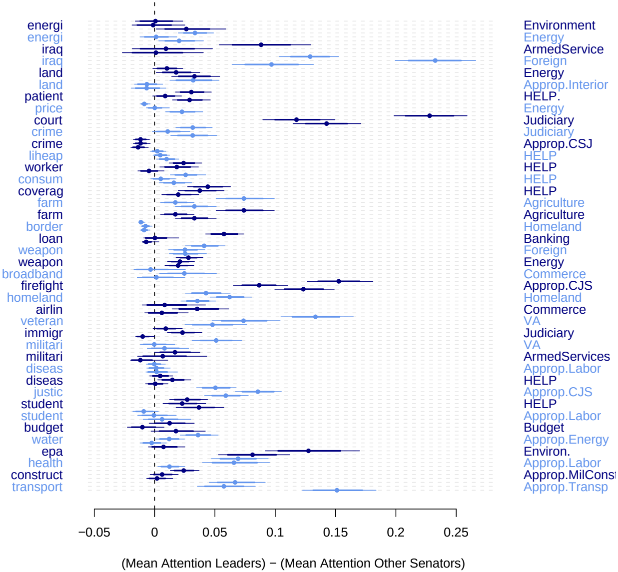
Home Style Measures, Convergent Validity



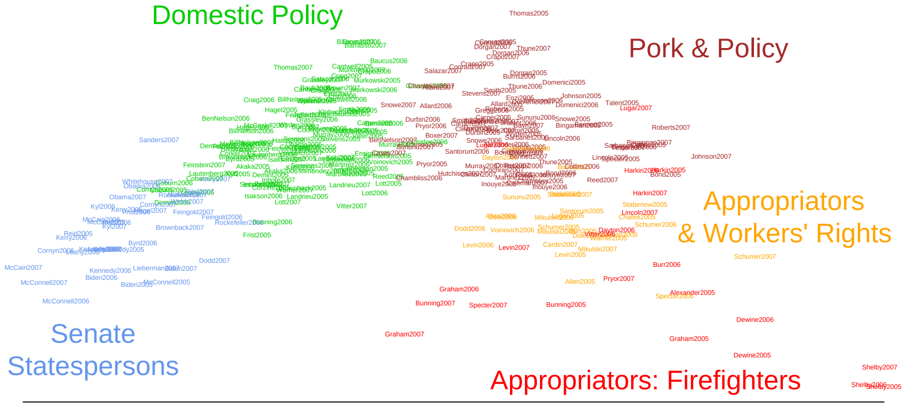
Discriminant Construct Validity



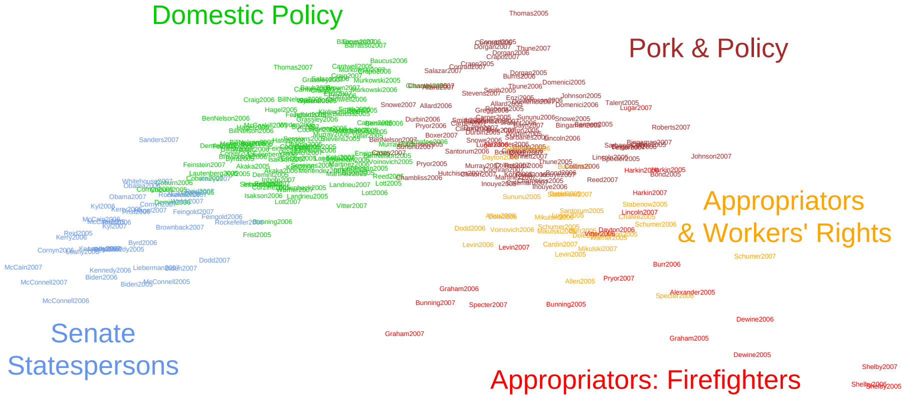
Predictive Validity



Hypothesis Validity



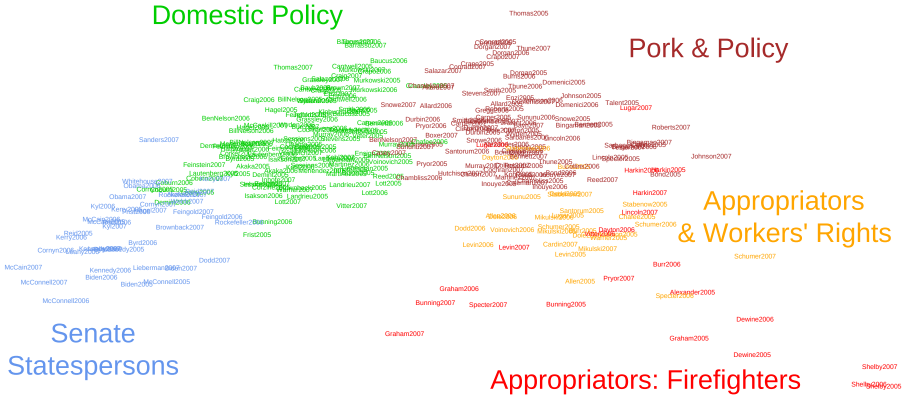
Hypothesis Validity



Senate Statesperson

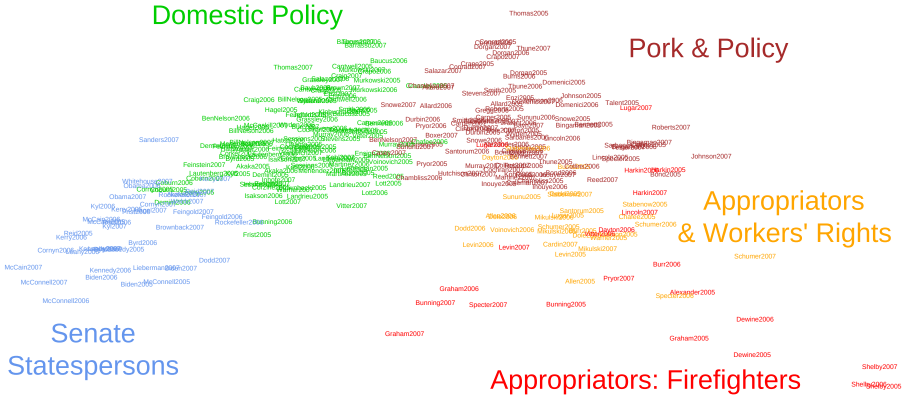
- Iraq War
- Intelligence
- Intl. Relations

Hypothesis Validity



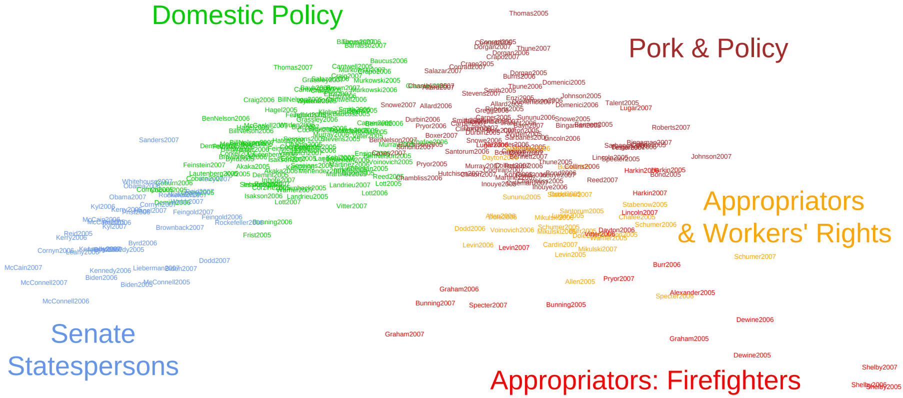
Senate Statesperson	Domestic Policy
- Iraq War	- Environment
- Intelligence	- Gas prices
- Intl. Relations	- DHS
	- Consumer

Hypothesis Validity



Senate Statesperson	Domestic Policy	Pork & Policy
- Iraq War	- Environment	- WRDA grants
- Intelligence	- Gas prices	- Farming
- Intl. Relations	- DHS	- Health Care
	- Consumer	- Education

Hypothesis Validity



Senate Statesperson
- Iraq War
- Intelligence
- Intl. Relations

Domestic Policy
- Environment
- Gas prices
- DHS
- Consumer

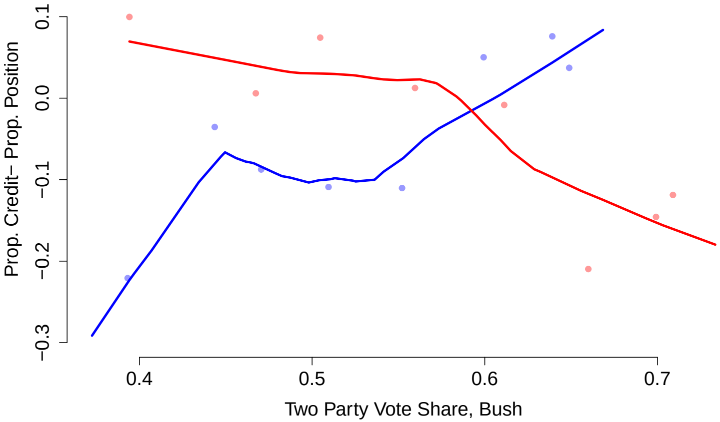
Pork & Policy
- WRDA grants
- Farming
- Health Care
- Education

Appropriators
- Fire Grants
- Airport Grants
- University Money

Hypothesis Validity

Why do senators adopt different styles?

District Fit



What are the right number of topics?

What are the right number of topics?

- Number of topics $\rightsquigarrow$ depends on task at hand

What are the right number of topics?

- Number of topics$\rightsquigarrow$ depends on task at hand
- Coarse$\rightsquigarrow$ broad comparisons, lose distinctions

What are the right number of topics?

- Number of topics$\rightsquigarrow$ depends on task at hand
- Coarse$\rightsquigarrow$ broad comparisons, lose distinctions
- Granular$\rightsquigarrow$ specific insights, lose broader picture

What are the right number of topics?

- Number of topics $\rightsquigarrow$ depends on task at hand
- Coarse $\rightsquigarrow$ broad comparisons, lose distinctions
- Granular $\rightsquigarrow$ specific insights, lose broader picture
- Hierarchy of topics $\rightsquigarrow$ Pachinko Allocation, Hierarchies of von-Mises Fisher Distributions

What are the right number of topics?

- Number of topics$\rightsquigarrow$ depends on task at hand
- Coarse$\rightsquigarrow$ broad comparisons, lose distinctions
- Granular$\rightsquigarrow$ specific insights, lose broader picture
- Hierarchy of topics$\rightsquigarrow$ Pachinko Allocation, Hierarchies of von-Mises Fisher Distributions

Blaydes, Grimmer, and McQueen [In Progress]$\rightsquigarrow$ estimate nested topics to explore the Mirros for Princes

The Mirrors Genre

26 Christian mirrors

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)
- The Sultan's Register of Laws (1632-1633 CE)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)
- The Sultan's Register of Laws (1632-1633 CE)

Work with translations

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)
- The Sultan's Register of Laws (1632-1633 CE)

Work with translations $\rightsquigarrow$ little evidence of selection

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)
- The Sultan's Register of Laws (1632-1633 CE)

Work with translations $\rightsquigarrow$ little evidence of selection

- Collect data on collection of 98 (51 Christian, 47 Islamic, some not translated)

The Mirrors Genre

26 Christian mirrors

- The Prince (1513 CE)
- Advice to Justinian (527 CE)
- The Adventures of Telemachus (1699 CE)

21 Islamic texts

- Advice on the Art of Governance (1612 CE)
- Kalila wa Dimna (748 CE)
- The Sultan's Register of Laws (1632-1633 CE)

Work with translations $\rightsquigarrow$ little evidence of selection

- Collect data on collection of 98 (51 Christian, 47 Islamic, some not translated)
- No difference on Year/Region

Preprocessing Texts

47 books

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
    - almighty, god

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
  - almighty, god
  - monarch, prince, king, ruler

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
  - almighty, god
  - monarch, prince, king, ruler
  - Lord $\neq$ lord

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
    - almighty, god
    - monarch, prince, king, ruler
    - Lord $\neq$ lord

Result: short segment j in book i is a count vector

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
  - almighty, god
  - monarch, prince, king, ruler
  - Lord $\neq$ lord

Result: short segment j in book i is a count vector

$$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ij2124})$$

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
  - almighty, god
  - monarch, prince, king, ruler
  - Lord $\neq$ lord

Result: short segment j in book i is a count vector

$$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ij2124})$$

We work with a normalized version of the documents,

Preprocessing Texts

47 books $\rightsquigarrow$ Each divided into paragraphs

Create feature space

- Bag of words, stem, discard punctuation, stop words
- Translate words left in Arabic (allah) and discard proper nouns
- Identified synonyms
  - almighty, god
  - monarch, prince, king, ruler
  - Lord $\neq$ lord

Result: short segment j in book i is a count vector

$$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ij2124})$$

We work with a normalized version of the documents,

$$\mathbf{x}_{ij}^{*} = \frac{\mathbf{x}_{ij}}{\sqrt{\mathbf{x}_{ij}^{'}\mathbf{x}_{ij}}}$$

Measuring Themes in the Mirrors

Model built around two hierarchies:

Measuring Themes in the Mirrors

Model built around two hierarchies:

1) Books $\rightsquigarrow$ paragraphs (Blei, Ng, Jordan 2003; Wallach, 2008; Quinn et al 2010; Grimmer 2010; Roberts et al 2014)

Measuring Themes in the Mirrors

Model built around two hierarchies:

1) Books $\rightsquigarrow$ paragraphs (Blei, Ng, Jordan 2003; Wallach, 2008; Quinn et al 2010; Grimmer 2010; Roberts et al 2014)
2) Coarse topics $\rightsquigarrow$ granular topics (Li and McCallum 2006; Gopal and Yang 2014)

Measuring Themes in the Mirrors

Estimate four quantities of interest

Measuring Themes in the Mirrors

Estimate four quantities of interest

1) Granular topics (60)

Measuring Themes in the Mirrors

Estimate four quantities of interest

1) Granular topics (60)
2) Coarse (broad) topics (3)

Measuring Themes in the Mirrors

Estimate four quantities of interest

1) Granular topics (60)
2) Coarse (broad) topics (3)
    - Each granular topic classified into one coarse topic

Measuring Themes in the Mirrors

Estimate four quantities of interest

1) Granular topics (60)
2) Coarse (broad) topics (3)
   - Each granular topic classified into one coarse topic
3) Each book i's $\textbf{themes}_i$

$$\textbf{themes}_i \quad = \quad (\text{theme}_{i1}, \text{theme}_{i2}, \ldots, \text{theme}_{i60})$$

Measuring Themes in the Mirrors

Estimate four quantities of interest

1) Granular topics (60)
2) Coarse (broad) topics (3)
    - Each granular topic classified into one coarse topic
3) Each book i's $\textbf{themes}_i$
4) Each short segment's granular (and coarse) topic

A Hierarchy of Topics

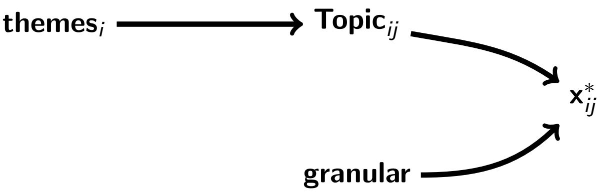
themes$_i$

A Hierarchy of Topics

$$\textbf{themes}_i \longrightarrow \textbf{Topic}_{ij}$$

$$\textbf{Topic}_{ij} \sim \text{Multinomial}(1, \textbf{themes}_i)$$

A Hierarchy of Topics



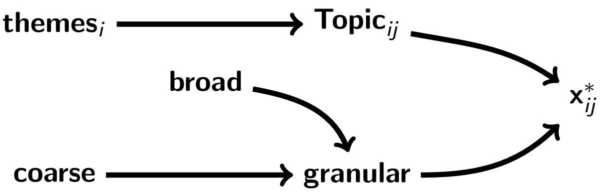
$$
\begin{aligned}
\textbf{Topic}_{ij} &\sim \text{Multinomial}(1, \textbf{themes}_i) \\
\mathbf{x}^*_{ij} | \text{Topic}_{ijk} = 1 &\sim \text{vMF}(\kappa, \textbf{granular}_k)
\end{aligned}
$$

A Hierarchy of Topics

$$
\begin{tikzcd}
\textbf{themes}_i \arrow[rr] & & \textbf{Topic}_{ij} \arrow[dr, bend left] & \\
& \textbf{broad} \arrow[dr, bend left] & & \mathbf{x}^*_{ij} \\
\textbf{coarse} \arrow[rr] & & \textbf{granular} \arrow[ur, bend right] &
\end{tikzcd}
$$

$$
\begin{aligned}
\textbf{Topic}_{ij} &\sim \text{Multinomial}(1, \textbf{themes}_i) \\
\mathbf{x}^*_{ij}|\text{Topic}_{ijk} = 1 &\sim \text{vMF}(\kappa, \textbf{granular}_k) \\
\textbf{broad}_k &\sim \text{Multinomial}(1, \textbf{Broad Theme Prior}) \\
\textbf{granular}_k|\text{broad}_{km} = 1 &\sim \text{vMF}(\kappa, \textbf{coarse}_m)
\end{aligned}
$$

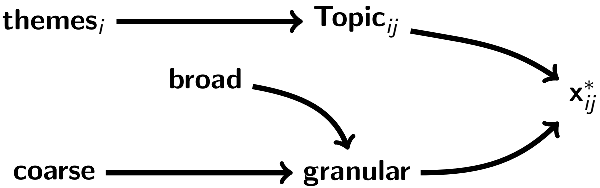
A Hierarchy of Topics



$$
\begin{aligned}
\textbf{Topic}_{ij} &\sim \text{Multinomial}(1, \textbf{themes}_i) \\
\mathbf{x}^*_{ij}|\text{Topic}_{ijk} = 1 &\sim \text{vMF}(\kappa, \textbf{granular}_k) \\
\textbf{broad}_k &\sim \text{Multinomial}(1, \textbf{Broad Theme Prior}) \\
\textbf{granular}_k|\text{broad}_{km} = 1 &\sim \text{vMF}(\kappa, \textbf{coarse}_m)
\end{aligned}
$$

Estimate model with Variational Approximation

A Hierarchy of Topics



$$\begin{aligned}
\textbf{Topic}_{ij} &\sim \text{Multinomial}(1, \textbf{themes}_i) \\
\mathbf{x}^*_{ij}|\text{Topic}_{ijk} = 1 &\sim \text{vMF}(\kappa, \textbf{granular}_k) \\
\textbf{broad}_k &\sim \text{Multinomial}(1, \textbf{Broad Theme Prior}) \\
\textbf{granular}_k|\text{broad}_{km} = 1 &\sim \text{vMF}(\kappa, \textbf{coarse}_m)
\end{aligned}$$

Estimate model with Variational Approximation
Model selection: automatic model fit, qualitative evaluation

Interpreting Unsupervised Models

Two approaches to labeling output

Interpreting Unsupervised Models

Two approaches to labeling output

1) Computational: identify discriminating words

Interpreting Unsupervised Models

Two approaches to labeling output

1) Computational: identify discriminating words
2) Manual: Segments classified to coarse, granular topics. Read, discuss, and label

Interpreting Unsupervised Models

Two approaches to labeling output

1) Computational: identify discriminating words
2) Manual: Segments classified to coarse, granular topics. Read, discuss, and label

Unsupervised models structure and guide our reading

Art of Rulership

Practices and ideals of political rule

Art of Rulership

Practices and ideals of political rule

king

Art of Rulership

Practices and ideals of political rule

king,princ

Art of Rulership

Practices and ideals of political rule

king,princ,citi

Art of Rulership

Practices and ideals of political rule

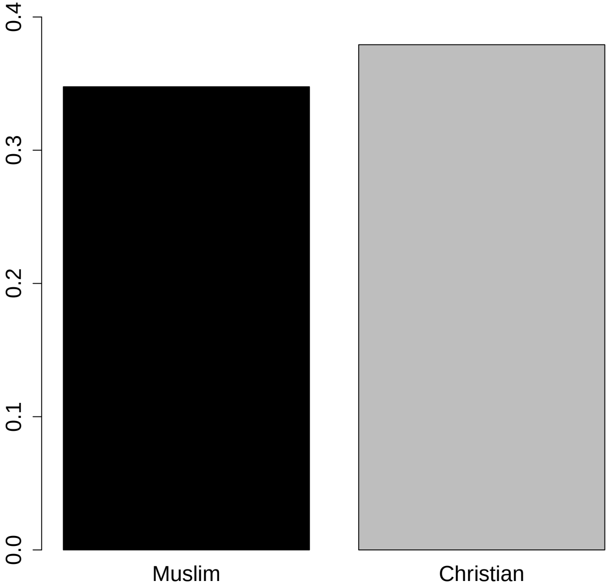
king,princ,citi,great,place,work,emperor,enemi,armi,letter

Art of Rulership

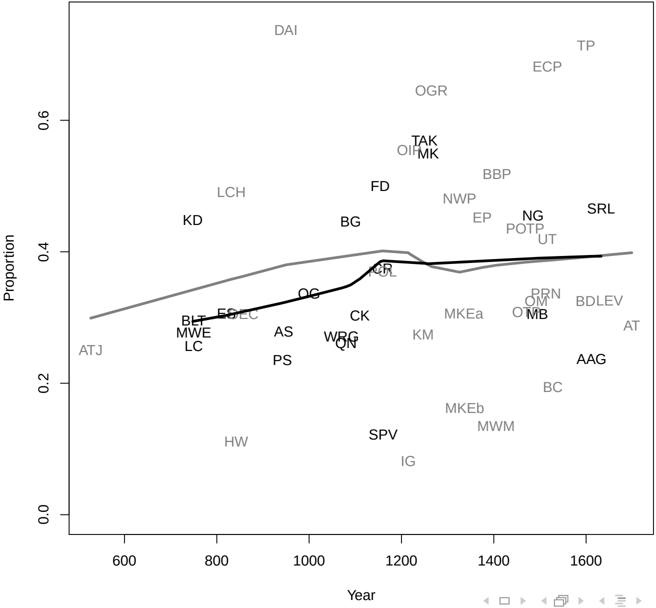
Practices and ideals of political rule

king,princ,citi,great,place,work,emperor,enemi,armi,letter

36.5% of paragraphs


0.0
0.1
0.2
0.3
0.4
Muslim
Christian

Coarse Topic 1



Religion and Virtue

Connection between religion, virtue, justice and political rule

Religion and Virtue

Connection between religion, virtue, justice and political rule

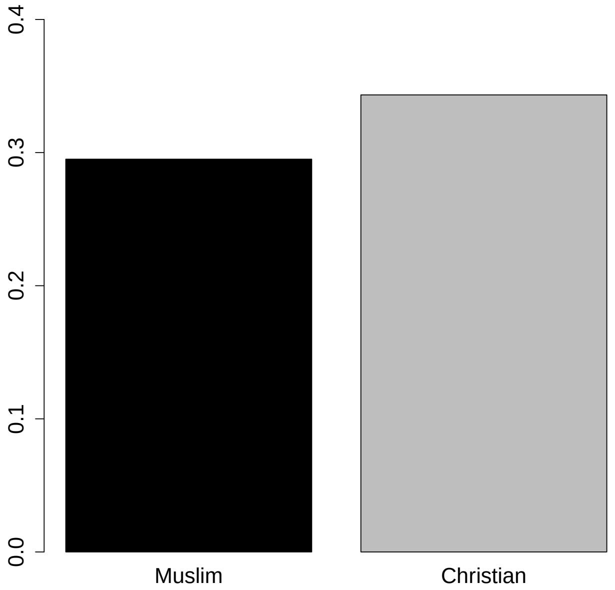
almighti,good,virtu,power,ruler,justic,prayer,rule,prophet,mena

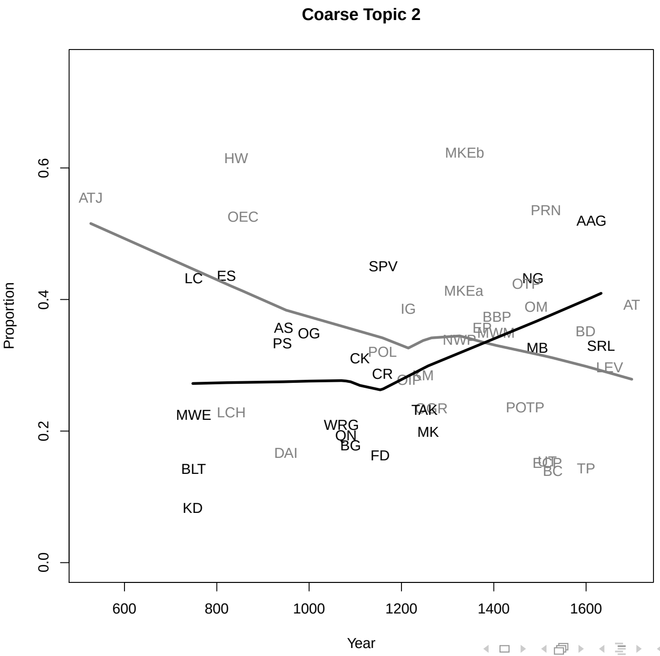
Religion and Virtue

Connection between religion, virtue, justice and political rule

almighti,good,virtu,power,ruler,justic,prayer,rule,prophet,mena

32.2% of pargraphs


0.0
0.1
0.2
0.3
0.4
Muslim
Christian


Coarse Topic 2
Proportion
Year
0.0
0.2
0.4
0.6
600
800
1000
1200
1400
1600
ATJ
HW
OEC
MKEb
PRN
AAG
SPV
LC
ES
NG
OTP
MKEa
IG
OM
AT
BBP
EP
MWM
AS
OG
PS
NWP
BD
SRL
MB
CK
POL
LEV
CR
OIP
KM
MWE
LCH
TAK
POTP
WRG
ON
BG
MK
DAI
FD
BLT
BC
TP
KD

Inner Life of the Ruler

Personal relationships, care for and practices of the self, and ultimate fate of the soul

Inner Life of the Ruler

Personal relationships, care for and practices of the self, and ultimate fate of the soul

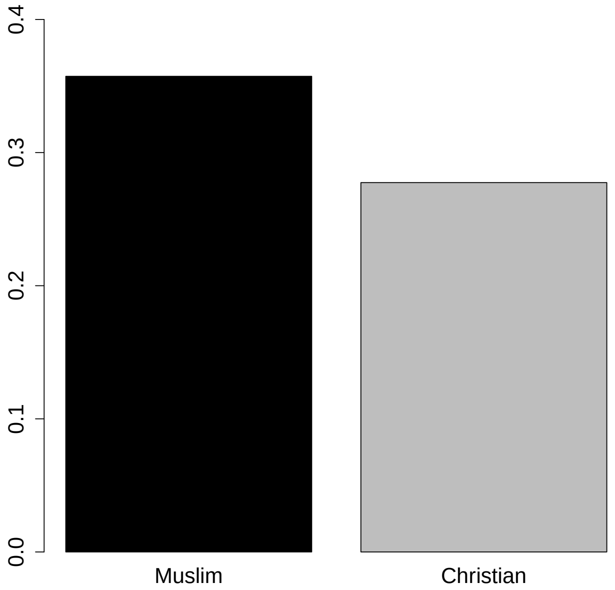
man,land,woman,know,bodi,eye,ladi,love,faculti,old

Inner Life of the Ruler

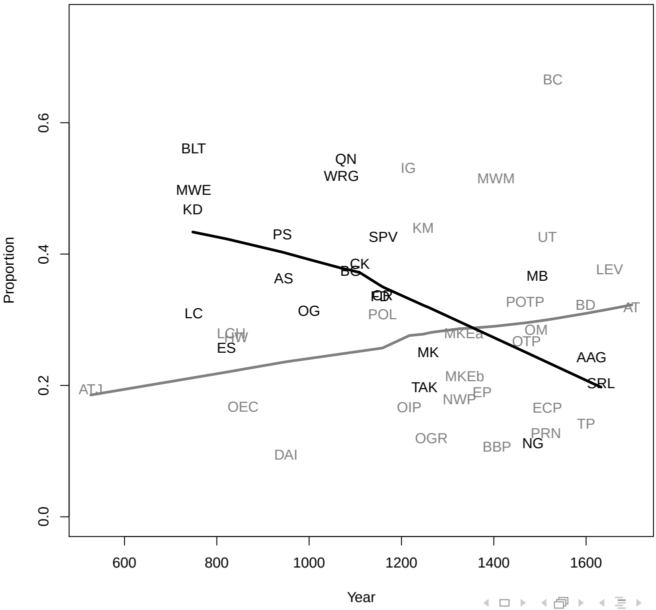
Personal relationships, care for and practices of the self, and ultimate fate of the soul

man,land,woman,know,bodi,eye,ladi,love,faculti,old

31.2% of paragraphs


0.0
0.1
0.2
0.3
0.4
Muslim
Christian

Coarse Topic 3

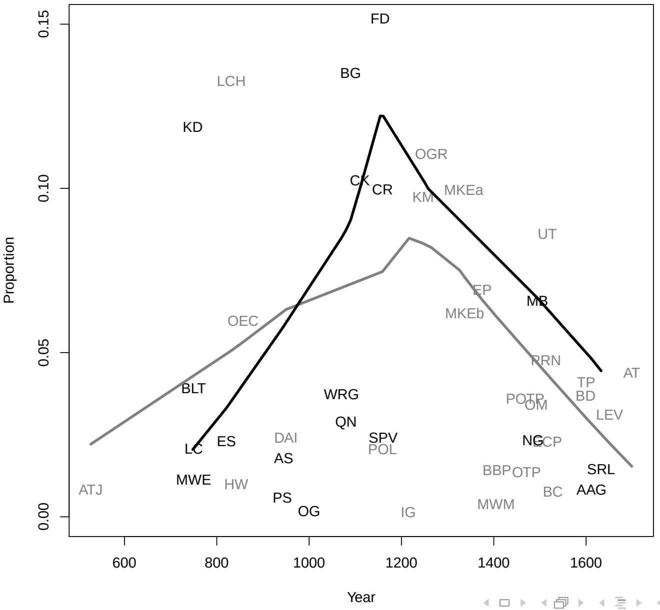


Granular: Best Practices for Ruling

king,princ,citi,great,place,work,emperor,enemi,armi,letter
king,kingdom,royal,minist,reign,father,court,majesti,presenc,war

6.2% of paragraphs

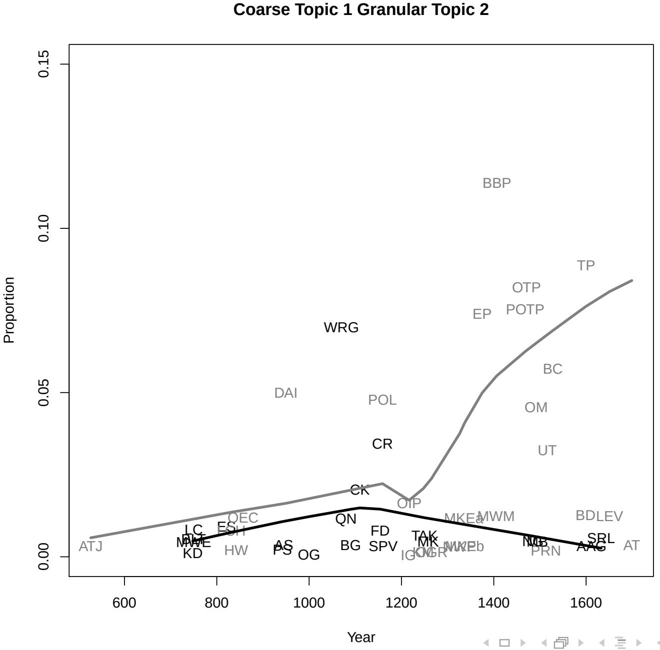
Coarse Topic 1 Granular Topic 1



Granular: Characteristics that distinguish Just Ruler from Tyrant

king,princ,citi,great,place,work,emperor,enemi,armi,letter
king,kingdom,royal,minist,reign,father,court,majesti,presenc,war
princ,good,peopl,christian,tyranni,war,mind,ought,state,public

3.1% of paragraphs

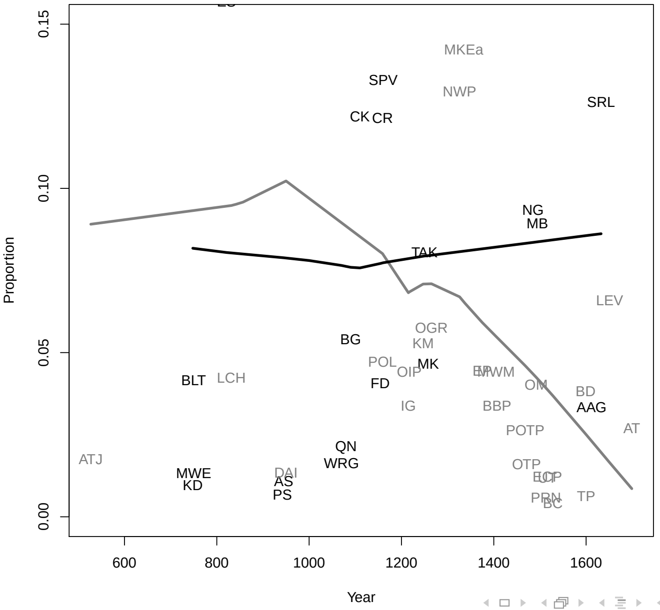

Coarse Topic 1 Granular Topic 2
Proportion
Year
0.00
0.05
0.10
0.15
600
800
1000
1200
1400
1600
BBP
TP
OTP
EP
POTP
WRG
BC
DAI
POL
OM
CR
UT
CK
OIP
QEC
MWM
BDLEV
QN
LC
FS
FD
TAK
SRL
MK
AS
BG
SPV
HW
PS
OG
IG
PRN
AAC
ATJ
AT
KD

Granular: Religious Virtues and Political Ideals

almighti,good,virtu,power,ruler,justic,prayer,rule,prophet,mena
almighti,bless,grant,peac,messeng,prophet,merci,holi,command,grace

6.9% of paragraphs

Coarse Topic 2 Granular Topic 1



Structural Topic Models

- Encode observed and unobserved meta data
- ?Improve substantive inferences

Next week:

1) Hanna Wallach on canonical topic models
2) Introduction to supervised learning

Work on your problem sets!

Appendix: Inference for both Models

Inference

Invariance in posterior makes it difficult (impossible) to approximate with sampling based methods (relabeling, aliasing problem).
Deterministic alternative: variational approximations.
Intuition: approximate posterior with simpler (still very general) approximating distribution.
Make approximation as "close" as possible

Approximate posterior with:

$$
\begin{aligned}
q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau}) &= q(\boldsymbol{\alpha})q(\boldsymbol{\beta})q(\boldsymbol{\theta})q(\boldsymbol{\sigma})q(\boldsymbol{\pi})q(\boldsymbol{\tau}) \\
&= q(\boldsymbol{\beta}) \prod_{k=1}^{K} q(\boldsymbol{\theta})_k \prod_{i=1}^{n} \prod_{t=2005}^{2007} \left[q(\boldsymbol{\sigma})_{it} q(\boldsymbol{\pi})_{it} \prod_{j=1}^{J} q(\boldsymbol{\tau})_{ijt} \right] \prod_{s=1}^{S} q(\boldsymbol{\alpha})_s
\end{aligned}
$$

Variational Approximation

Optimization goal:

- Minimize the Kullback-Leibler divergence between approximating distribution q and true posterior p
  - KL-divergence is a functional: takes functions as an input, returns a positive scalar
  - Measures "divergence" between two measures
- Use calculus of variations and theory from exponential models to derive iterative algorithm
- See "An Introduce to Bayesian Inference via Variational Approximations" for extended introduction (Grimmer, 2011)

Minimize KL-divergence by Maximizing a Lower Bound

$$\log p(\boldsymbol{Y}) = \log \sum_{\boldsymbol{\sigma}} \sum_{\boldsymbol{\tau}} \iiiint p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{Y}) d\boldsymbol{\theta} d\boldsymbol{\alpha} d\boldsymbol{\beta} d\boldsymbol{\pi}$$

$$\log p(\boldsymbol{Y}) = \log \sum_{\boldsymbol{\sigma}} \sum_{\boldsymbol{\tau}} \iiiint p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{Y}) \frac{q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau})}{q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau})} d\boldsymbol{\theta} d\boldsymbol{\alpha} d\boldsymbol{\beta} d\boldsymbol{\pi}$$

$$\log p(\boldsymbol{Y}) \geq \underbrace{\sum_{\boldsymbol{\sigma}} \sum_{\boldsymbol{\tau}} \iiiint q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau}) \log \frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau} | \boldsymbol{Y})}{q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\tau})} d\boldsymbol{\theta} d\boldsymbol{\alpha} d\boldsymbol{\beta} d\boldsymbol{\pi}}_{\mathcal{L}(q)}.$$

Minimize KL-divergence by Maximizing a Lower Bound

$$\log p(\boldsymbol{Y}) = \mathcal{L}(q) + \text{KL}(q||p)$$

Minimize KL-divergence by Maximizing a Lower Bound

$$
\begin{aligned}
\log p(\boldsymbol{Y}) &= \mathcal{L}(q) + \mathrm{KL}(q||p) \\
\underbrace{\log p(\boldsymbol{Y})}_{\text{fixed number}} &= \mathcal{L}(q) + \underbrace{\mathrm{KL}(q||p)}_{\text{Positive}}
\end{aligned}
$$

Minimize KL-divergence by Maximizing a Lower Bound

$$
\begin{aligned}
\log p(\boldsymbol{Y}) &= \mathcal{L}(q) + \mathrm{KL}(q||p) \\
\underbrace{\log p(\boldsymbol{Y})}_{\text{fixed number}} &= \mathcal{L}(q) + \underbrace{\mathrm{KL}(q||p)}_{\text{Positive}}
\end{aligned}
$$

If $\mathcal{L}(q)$ get bigger, $\mathrm{KL}(q||p)$ get smaller. $\Rightarrow$

Minimize KL-divergence by Maximizing a Lower Bound

$$
\begin{aligned}
\log p(\boldsymbol{Y}) &= \mathcal{L}(q) + \text{KL}(q||p) \\
\underbrace{\log p(\boldsymbol{Y})}_{\text{fixed number}} &= \mathcal{L}(q) + \underbrace{\text{KL}(q||p)}_{\text{Positive}}
\end{aligned}
$$

If $\mathcal{L}(q)$ get bigger, $\text{KL}(q||p)$ get smaller. $\Rightarrow$
If $\mathcal{L}(q)$ is at a maximum $\text{KL}(q||p)$ is at a minimum (duals).

Maximizing $\mathcal{L}(q)$

Goal: choose q to maximize $\mathcal{L}(q)$.

$$q(\boldsymbol{\alpha})^{\text{old}}, q(\boldsymbol{\beta})^{\text{old}}, q(\boldsymbol{\theta})^{\text{old}}, q(\boldsymbol{\sigma})^{\text{old}}, q(\boldsymbol{\pi})^{\text{old}}, q(\boldsymbol{\tau})^{\text{old}}.$$

Iterative Algorithm:
Choose, $q(\boldsymbol{\pi})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\theta})^{\text{old}}, q(\boldsymbol{\tau})^{\text{old}}, q(\boldsymbol{\alpha})^{\text{old}}, q(\boldsymbol{\beta})^{\text{old}}, q(\boldsymbol{\sigma})^{\text{old}}$ constant.
Choose, $q(\boldsymbol{\theta})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\pi})^{\text{new}}, q(\boldsymbol{\tau})^{\text{old}}, q(\boldsymbol{\alpha})^{\text{old}}, q(\boldsymbol{\beta})^{\text{old}}, q(\boldsymbol{\sigma})^{\text{old}}$ constant.
Choose, $q(\boldsymbol{\tau})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\theta})^{\text{new}}, q(\boldsymbol{\pi})^{\text{new}}, q(\boldsymbol{\alpha})^{\text{old}}, q(\boldsymbol{\beta})^{\text{old}}, q(\boldsymbol{\sigma})^{\text{old}}$ constant.
Choose, $q(\boldsymbol{\alpha})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\theta})^{\text{new}}, q(\boldsymbol{\tau})^{\text{new}}, q(\boldsymbol{\pi})^{\text{new}}, q(\boldsymbol{\beta})^{\text{old}}, q(\boldsymbol{\sigma})^{\text{old}}$ constant.
Choose, $q(\boldsymbol{\beta})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\theta})^{\text{new}}, q(\boldsymbol{\tau})^{\text{new}}, q(\boldsymbol{\pi})^{\text{new}}, q(\boldsymbol{\alpha})^{\text{new}}, q(\boldsymbol{\sigma})^{\text{old}}$ constant.
Choose, $q(\boldsymbol{\sigma})^{\text{new}}$ to max $\mathcal{L}(q)$–holding $q(\boldsymbol{\theta})^{\text{new}}, q(\boldsymbol{\tau})^{\text{new}}, q(\boldsymbol{\pi})^{\text{new}}, q(\boldsymbol{\alpha})^{\text{new}}, q(\boldsymbol{\beta})^{\text{new}}$ constant.

Example for $q(\boldsymbol{\pi})^{\text{new}}$

$$\mathcal{L}(q) = \int q(\boldsymbol{\pi})^{\text{new}} \underbrace{\left\{ \sum_{\boldsymbol{\sigma}} \sum_{\boldsymbol{\tau}} \iiint \log p(\boldsymbol{Y}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\tau}) q(\boldsymbol{\sigma})^{\text{old}} q(\boldsymbol{\tau})^{\text{old}} q(\boldsymbol{\theta})^{\text{old}} q(\boldsymbol{\alpha})^{\text{old}} q(\boldsymbol{\beta})^{\text{old}} d\boldsymbol{\theta} d\boldsymbol{\alpha} d\boldsymbol{\beta} \right\}}_{\mathrm{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\beta}}[\log p(\boldsymbol{Y},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\beta},\boldsymbol{\sigma})]} d\boldsymbol{\pi}$$
$$- q(\boldsymbol{\pi})^{\text{new}} \log q(\boldsymbol{\pi})^{\text{new}} + \text{constants}$$

Define

$$\log \tilde{p}(\boldsymbol{\pi}) = \mathrm{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\beta}}[\log p(\boldsymbol{Y}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\sigma})] + \text{constants}$$

Example for $q(\boldsymbol{\pi})^{\text{new}}$

Substituting $\log \tilde{p}(\boldsymbol{\pi})$,

$$\begin{aligned} &= \int q(\boldsymbol{\pi})^{\text{new}} \log \left(\frac{\tilde{p}(\boldsymbol{\pi})}{q(\boldsymbol{\pi})^{\text{new}}} \right) d\boldsymbol{\pi} \\ &= -\text{KL}(q(\boldsymbol{\pi})^{\text{new}} || \tilde{p}(\boldsymbol{\pi})) \end{aligned}$$

$\Rightarrow$ At a maximum when $q(\boldsymbol{\pi})^{\text{new}} = \tilde{p}(\boldsymbol{\pi})$.
Equivalently,

$$\begin{aligned} \log q(\boldsymbol{\pi})^{\text{new}} &= \log \tilde{p}(\boldsymbol{\pi}) \\ &= \text{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\beta}}[\log p(\boldsymbol{Y}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\sigma})] + \text{constants} \end{aligned}$$

Or,

$$q(\boldsymbol{\pi})^{\text{new}} = \frac{\exp(\text{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\beta}}[\log p(\boldsymbol{Y}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\sigma})])}{\int \exp(\text{E}_{\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\beta}}[\log p(\boldsymbol{Y}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\sigma})]) d\boldsymbol{\pi}}$$

To maximize $\mathcal{L}(q)$ we use the following iterative algorithm

$$
\begin{aligned}
q(\boldsymbol{\sigma})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right) \\
q(\boldsymbol{\tau})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\sigma},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right) \\
q(\boldsymbol{\theta})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right) \\
q(\boldsymbol{\alpha})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\pi}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right) \\
q(\boldsymbol{\beta})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\pi}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right) \\
q(\boldsymbol{\pi})^{\text{new}} & \propto \exp\left(\mathrm{E}_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta}}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{Y})]\right)
\end{aligned}
$$

Update for $q(\boldsymbol{\sigma})_{it}$

$q(\boldsymbol{\sigma})_{it}$ is a Multinomial$(1, \boldsymbol{c}_{it})$ distribution, with typical parameter c_{its}

$$\boldsymbol{c}_{its} \quad \propto \quad \exp\left\{ \mathrm{E}[\log \beta_s] + \log \Gamma(\sum_{k=1}^{K} \alpha_{ks}) - \sum_{k=1}^{K} \log \Gamma(\alpha_{ks}) + \sum_{k=1}^{K} (\alpha_{ks} - 1) \mathrm{E}[\log \pi_{itk}] \right\}.$$

Update for $q(\boldsymbol{\tau})_{ijt}$

$q(\boldsymbol{\tau})_{ijt}$ is a Multinomial$(1, \boldsymbol{r}_{ijt})$ distribution with typical parameter,

$$r_{ijtk} \quad \propto \quad \exp\left\{ \mathrm{E}[\log \pi_{itk}] + \sum_{w=1}^{W} y_{ijtw} \mathrm{E}[\log \theta_{kw}] \right\}.$$

Update for $q(\boldsymbol{\pi})_{it}$

$q(\boldsymbol{\pi})_{it}$ is a Dirichlet(γ_{it}) distribution, with typical parameter γ_{itk} equal to

$$\gamma_{itk} = \sum_{s=1}^{S} c_{its}\alpha^{*}_{sk} + \sum_{j=1}^{D_{it}} r_{ijtk}$$

Update for $q(\boldsymbol{\theta})_k$

$q(\boldsymbol{\theta})_k$ is a Dirichlet($\boldsymbol{\eta}_k$) distribution, with typical parameter equal to,

$$\eta_{kw} = \lambda_w + \sum_{i=1}^{n} \sum_{t=2005}^{2007} \sum_{j=1}^{D_{it}} r_{itjk} y_{itw}$$

Update for $q(\boldsymbol{\beta})$

$q(\boldsymbol{\beta})$ is a Dirichlet(ϕ) distribution, with typical parameter ϕ_s equal to,

$$\phi_s = 1 + \sum_{i=1}^{n} \sum_{t=2005}^{2007} c_{its}$$

Completing $q(\boldsymbol{\sigma})_{it}$ and $q(\boldsymbol{\tau})_{ijt}$

Finishing $q(\boldsymbol{\sigma})_{it}$:

- $\text{E}[\log \beta_s] = \Psi(\phi_s) - \Psi(\sum_{z=1}^{S} \phi_z)$ where $\Psi(\cdot)$ is the digamma function (the derivative of the gamma function)
- $\text{E}[\log \pi_{itk}] = \Psi(\gamma_{itk}) - \Psi(\sum_{z=1}^{K} \gamma_{itz})$

Finishing $q(\boldsymbol{\tau})_{ijt}$

- $\text{E}[\log \theta_{kw}] = \Psi(\eta_{kw}) - \Psi(\sum_{z=1}^{w} \eta_{kz})$.

Update Steps for $\boldsymbol{\alpha}_s$

(Newton-Raphson, Minka 2000)

- Define $N_s = \sum_{i=1}^{n} \sum_{t=2005}^{2007} c_{its}$.

Differentiating with respect to α_{ks} shows that

$$\frac{\partial \log q(\boldsymbol{\alpha})_k^{\text{new}}}{\partial \alpha_{ks}} = -\frac{1}{\lambda} + N_s \Psi(\sum_{k=1}^{K} \alpha_{ks}) - N_s \Psi(\alpha_{ks}) + \sum_{i=1}^{n} \sum_{t=2005}^{2007} c_{its} \frac{\left(\Psi(\gamma_{itk}) - \Psi(\sum_{z=1}^{K} \gamma_{itz})\right)}{Ns}$$

- Call Gradient $\frac{\partial \log q(\boldsymbol{\alpha})_k^{\text{new}}}{\partial \boldsymbol{\alpha}_k}$.
- Define H as the Hessian (matrix of second derivatives).
- Diagonal element $h_{jj} = N_s \Psi^{'}(\sum_{k=1}^{K} \alpha_{ks}) - N_s \Psi^{'}(\alpha_{js})$ where $\Psi^{'}(\cdot)$ is the trigamma function
- Off-diagonal element $(a \neq b)$ $h_{ab} = N_z \Psi^{'}(\sum_{k=1}^{K} \alpha_{ks})$.

For each s we iterate,

$$\boldsymbol{\alpha}_s^{\text{new}} = \boldsymbol{\alpha}_s^{\text{old}} - \mathrm{H}^{-1} \frac{\partial \log q(\boldsymbol{\alpha})_k^{\text{new}}}{\partial \boldsymbol{\alpha}_k}$$

until convergence

Initialize $\boldsymbol{\gamma}_{it}^{\text{old}}$ (for all i and t), $\boldsymbol{\eta}_k^{\text{old}}$ (for all k), $\boldsymbol{\phi}^{\text{old}}$, $\boldsymbol{\alpha}_s^{\text{old}}$ (for all s).
Do until convergence in lower-bound.
- for all i, t, j and k, set

$$r_{ijtk}^{\text{new}} \propto \exp\left(\Psi(\gamma_{itk}^{\text{old}}) - \Psi(\sum_{z=1}^{K} \gamma_{itz}^{\text{old}}) + \sum_{w=1}^{W} y_{ijtw}\left[\Psi(\eta_{kw}^{\text{old}}) - \Psi(\sum_{z=1}^{W} \eta_{kz}^{\text{old}})\right]\right)$$

-for all i,t, and s set

$$\begin{aligned} c_{its}^{\text{new}} \propto \exp(\Psi(\phi_s^{\text{old}}) - \Psi(\sum_{z=1}^{S} \phi_s^{\text{old}}) + \log\Gamma(\sum_{k=1}^{K} \alpha_{ks}^{\text{old}}) - \sum_{k=1}^{K} \log\Gamma(\alpha_{ks}^{\text{old}}) \\ + \sum_{k=1}^{K}(\alpha_{ks}^{\text{old}} - 1)[\Psi(\gamma_{itk}^{\text{old}}) - \Psi(\sum_{z=1}^{K} \gamma_{itz}^{\text{old}})]) \end{aligned}$$

- for all i, t , and k set

$$\gamma_{itk}^{\text{new}} = \sum_{s=1}^{S} c_{its}^{\text{new}} \alpha_{ks}^{\text{old}} + \sum_{j=1}^{D_{it}} r_{ijtk}^{\text{new}}$$

- for all k and w set

$$\eta_{kw}^{\text{new}} = \lambda_w + \sum_{i=1}^{n} \sum_{t=2005}^{2007} \sum_{j=1}^{D_{it}} r_{ijtk}^{\text{new}} y_{ijtw}$$

-for all s set

$$\phi_s^{\text{new}} = 1 + \sum_{i=1}^{n} \sum_{t=2005}^{2007} c_{its}^{\text{new}}$$

- For all s obtain $\boldsymbol{\alpha}_s^{\text{new}}$ using Newton-Raphson algorithm.
- Evaluate lower-bound.

If converged:
Return posterior approximation.

$<$ / Variational Approximation $>$

< Model Selection >

Number of Topics

1) Substantive search (about 40-50)
2) 10-fold cross-validation. Loss function, approximate predictive posterior

$$p(\hat{\boldsymbol{y}}|\boldsymbol{Y}) \approx \sum_{\hat{\boldsymbol{\tau}}} \int\int p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{\tau}}, \boldsymbol{\theta}) p(\hat{\boldsymbol{\tau}}|\boldsymbol{\pi}) q(\boldsymbol{\theta}|\boldsymbol{\eta}) q(\boldsymbol{\pi}) d\boldsymbol{\theta} d\boldsymbol{\pi}$$

3) Convergence with Nonparametric Bayesian model (Dirichlet process prior)

All converge on about 44 topics

Number of Styles

$$\text{BIC} = 2 \log p(\boldsymbol{Y})$$

1) $\text{BIC} \approx 2(\mathcal{L}(q) + \log K! + \log S!) - (K \times S)(n)$
2) $\text{BIC} \approx 2 \log p(\boldsymbol{Y} | \bar{\boldsymbol{\pi}}, \bar{\theta}, \bar{\boldsymbol{\tau}}) - (K \times S)(n)$

< / Model Selection >

Hierarchy of Topics

Posterior Distribution

$$p(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\tau} | \boldsymbol{X}) \propto \prod_{m=1}^{M} c(\kappa) \exp\left(\kappa \boldsymbol{\eta}_m^{'} \frac{\mathbf{1}}{\sqrt{\boldsymbol{J}}}\right) \times \prod_{m=1}^{M} \prod_{k=1}^{K} \left[\beta_m c(\kappa) \exp\left(\kappa \boldsymbol{\mu}_k^{'} \boldsymbol{\eta}_m\right)\right]^{\sigma_{mk}} \prod_{k=1}^{K} \exp(-\alpha_k) \times$$

$$\prod_{i=1}^{48} \left[\frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} (\alpha_k)} \prod_{k=1}^{K} \pi_{ik}^{\alpha_k - 1} \times \prod_{j=1}^{D_i} \prod_{k=1}^{K} \left[\pi_{ik} c(\kappa) \exp(\kappa \boldsymbol{x}_{ij}^{*} \boldsymbol{\mu}_k\right]^{\tau_{ijk}} \right] \quad (0.1)$$

Which we approximate with:

$$q(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\tau}) = q(\boldsymbol{\alpha}) q(\boldsymbol{\pi}) q(\boldsymbol{\eta}) q(\boldsymbol{\beta}) q(\boldsymbol{\sigma}) q(\boldsymbol{\mu}) q(\boldsymbol{\tau}) \quad (0.2)$$

$$= q(\boldsymbol{\alpha}) \prod_{i=1}^{48} q(\boldsymbol{\pi})_i \prod_{m=1}^{M} q(\boldsymbol{\eta})_m q(\boldsymbol{\beta}) \prod_{k=1}^{K} q(\boldsymbol{\sigma})_k \prod_{k=1}^{K} q(\boldsymbol{\mu})_k \prod_{i=1}^{48} \prod_{j=1}^{D_i} q(\boldsymbol{\tau})_{ij}$$

Update for $q(\boldsymbol{\sigma})_k$

$q(\boldsymbol{\sigma})_k$ is a Multinomial(1, $\boldsymbol{c}_k$) where typical element c_{mk} is equal to

$$c_{mk} \quad \propto \quad \exp\left(\mathrm{E}[\log \beta_m] + \mathrm{E}[\kappa \boldsymbol{\mu}_k \boldsymbol{\eta}_m]\right).$$

We will complete the update step when we have determined the remaining forms of the distribution

Update for $q(\boldsymbol{\tau})_{ij}$

$q(\boldsymbol{\tau})_{ij}$ is a Multinomial(1, $\boldsymbol{r}_{ij}$, with typical element of r_{ijk} equal to

$$r_{ijk} \quad \propto \quad \exp\left(\mathrm{E}[\log \pi_{ik}] + \mathrm{E}[\kappa \boldsymbol{y}_{ij}^{*} \boldsymbol{\mu}_k]\right)$$

Again, as we complete the parametric forms of the other update steps we can complete this update equation.

Update for $q(\boldsymbol{\pi})_i$

$q(\boldsymbol{\pi})_i$ is a Dirichlet($\boldsymbol{\gamma}_i$) distribution, where typical element γ_{ik} is equal to

$$\gamma_{ik} = \alpha_k + \sum_{j=1}^{D_i} r_{ijk}$$

Update for $q(\boldsymbol{\beta})$

$q(\boldsymbol{\beta})$ is a Dirichlet($\boldsymbol{\phi}$) distribution with typical parameter ϕ_m equal to

$$\phi_m = 1 + \sum_{k=1}^{K} c_{mk}$$

Update for $q(\boldsymbol{\eta})_m$

Given the complications of taking expectations with the vMF distribution, we instead provide maximization steps for the vMF parameters. To obtain the form of the updates we follow the derivation outlined in Banerjee et al (2005). To do this, we take the log of the posterior distribution and identify the parameters that depend upon $\boldsymbol{\eta}_m$.

$$\log(p(\boldsymbol{\eta}_m) \quad = \quad \sum_{k=1}^{K} c_{km}\kappa\boldsymbol{\mu}_k\boldsymbol{\eta}_m + \kappa\boldsymbol{\eta}_m\frac{\mathbf{1}}{\sqrt{\boldsymbol{J}}} + \text{constants}$$

Update for $q(\boldsymbol{\eta})_m$

To set up the constrained optimization we also introduce the Langragian λ, with the constraint that $\boldsymbol{\eta}_m^{'}\boldsymbol{\eta}_m = 1$,

$$\log(p(\boldsymbol{\eta}_m) \quad \propto \quad \sum_{k=1}^{K} c_{km}\kappa\boldsymbol{\mu}_k\boldsymbol{\eta}_m + \kappa\boldsymbol{\eta}_m\frac{\mathbf{1}}{\sqrt{\boldsymbol{J}}} - \lambda(\boldsymbol{\eta}_m^{'}\boldsymbol{\eta}_m - 1).$$

Differentiating with respect to $\boldsymbol{\eta}_m$, setting equal to zero and solving yields

$$\frac{\kappa\left(\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{\boldsymbol{J}}}\right)}{2\lambda} \quad = \quad \boldsymbol{\eta}_m \qquad (0.3)$$

Update for $q(\boldsymbol{\eta})_m$

If we differentiate with respect to λ and solve we see that $\boldsymbol{\eta}_m^{'}\boldsymbol{\eta}_m = 1$ or that $||\boldsymbol{\eta}_m^{'}\boldsymbol{\eta}_m|| = 1$. Substituting this into Equation 0.3 we have,

$$
\begin{aligned}
\frac{\kappa}{2\lambda}\left(\left(\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{J}}\right)^{'}\left(\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{J}}\right)\right)^{1/2} &= 1 \\
\frac{\kappa||\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{J}}||}{2} &= \lambda
\end{aligned}
$$

Doing a final substitution we have

$$
\boldsymbol{\eta}_m^{*} = \frac{\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{J}}}{||\sum_{k=1}^{K} c_{mk}\boldsymbol{\mu}_k + \frac{\mathbf{1}}{\sqrt{J}}||}
$$

Update for $q(\boldsymbol{\mu})_k$

Following a very similar set of derivations, the update step for $\boldsymbol{\mu}_k$ is

$$\boldsymbol{\mu}_k^* = \frac{\sum_{i=1}^{48}\sum_{j=1}^{D_i} r_{ijk}\boldsymbol{x}_{ij}^* + \sum_{m=1}^{M} c_{mk}\boldsymbol{\eta}_m^*}{||\sum_{i=1}^{48}\sum_{j=1}^{D_i} r_{ijk}\boldsymbol{x}_{ij}^* + \sum_{m=1}^{M} c_{mk}\boldsymbol{\eta}_m^*||}$$

Completing updates for $q(\boldsymbol{\sigma})_k$ and $q(\boldsymbol{\tau})_{ij}$

Given the forms $\text{E}[\log \beta_m] = \Psi(\phi_m) - \Psi(\sum_{m=1}^{M} \phi_m)$ and $\text{E}[\log \pi_{ik}] = \Psi(\gamma_{ik}) - \Psi(\sum_{k=1}^{K} \gamma_{ik})$ where $\Psi(\cdot)$ is the Digamma function.

Update for q(α)

A closed form update for the α parameters is unavailable. So we use the Newton-Raphson algorithm outlined in Minka (2000) and Blei, Ng, and Jordan (2003).